April 5th, 2024

INF2178

Shion Guha

Benjamin Rabishaw

INF2178 Technical Assignment 4:

Mixed ANOVAs

I. Introduction

Dementia is a term for a collection of brain disorders affecting memory, cognition, emotion, and behaviour (ADI, n.d.). It is estimated that the overall cost of dementia management and care exceeds 1.3 trillion USD (ADI, n.d.), and by 2050, the world's population with dementia is estimated to grow beyond 152 million people (ADI, 2019). Dementia is most commonly caused by Alzheimer's disease, but can also be caused by vascular diseases and others (ADI, n.d.). A person living with dementia may encounter: difficulty performing familiar tasks, including expressing themselves with language; spatial and temporal disorientation; rapid changes in mood; problems with concentration or future planning; and memory loss (ADI, n.d.).

An important aspect of dementia is that the direct symptoms themselves are not the only unhealthy outcomes. With dementia especially, social stigma and lack of robust care resources are significant challenges (ADI, 2019). Quantitative work on public attitudes toward dementia suggest that one in four people believe dementia cannot be mitigated; more than 60% of healthcare workers believe, incorrectly, that dementia is a necessary part of aging; and up to 60% of demented people report being treated unfairly in their relationships (ADI, 2019). Moreover, the social support networks and caregivers for people with dementia report challenges to do with their own mental, physical, and social health as a downstream impact of supporting a demented person (Vliet et al., 2010). Dementia care and outcomes do have a socioeconomic component, with dementia patients from low and middle-income areas reporting higher social stigma and weaker support than their high-income counterparts (ADI, 2019). Considering the challenges involved with caring for people with dementia, the increased number of people with dementia, and the relationship between socioeconomic status, robustness and support, and outcomes for patients, it may be valuable to investigate the relationships between markers of dementia and socioeconomic status.

Using a dataset of MRI data for people with dementia and people without, this technical assignment will investigate two main research questions:

1. Does the relationship between socioeconomic status and cognitive function differ between people with and without dementia?
2. What effect does socioeconomic status have on cognitive function over repeated visits?

II. The Data and Data Cleaning

The data was retrieved from the INF2178 Quercus module as a CSV or comma-separated value file format. The data originates from the data science competition platform Kaggle, where the publication is called “MRI and Alzheimers: Magnetic Resonance Imaging Comparisons of Demented and Nondemented Adults,” uploaded by Jacob Boysen under a Public Domain Creative Commons license. The name of the Quercus CSV is “INF2178_A4_data.csv” and has a length of 294, with columns including:

- Group, meaning Demented or Nondemented, with a Converted status for subjects who became demented over the course of the study;
- SES, meaning socioeconomic quintile;
- MMSE, or mini-mental state examination score, the result of a 30-point questionnaire to identify MCI, or mild cognitive impairment, where a score of 24 or above indicates normal cognition;

Because all the subjects are right-handed, data cleaning involved removing a column for handedness, as well as removing rows for which there was no SES data, in line with this assignment’s research questions.

III. Power Analysis

The first step of the investigation, which must take place *a priori*, is the calculation of the appropriate sample size and creation of a statistical power analysis plot for t-tests. Given a desired power of 0.91, an alpha of 0.05, and an effect size of 0.7, the calculated necessary sample size is at least 45, with the plot below (Figure 1, below):

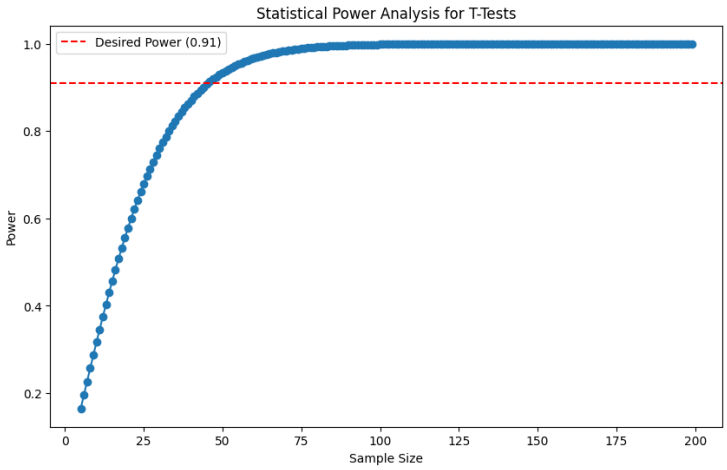

Figure 1

As the cleaned dataset has a total length of 279, including 142 unique subjects, the dataset appears to satisfy and exceed the minimum sample size indicated by the power analysis and curve.

IV. Research Question #1: Between-Subjects

Exploratory data analysis for RQ1 included a boxplot showing MMSE scores grouped by dementia status (Figure 2, below):

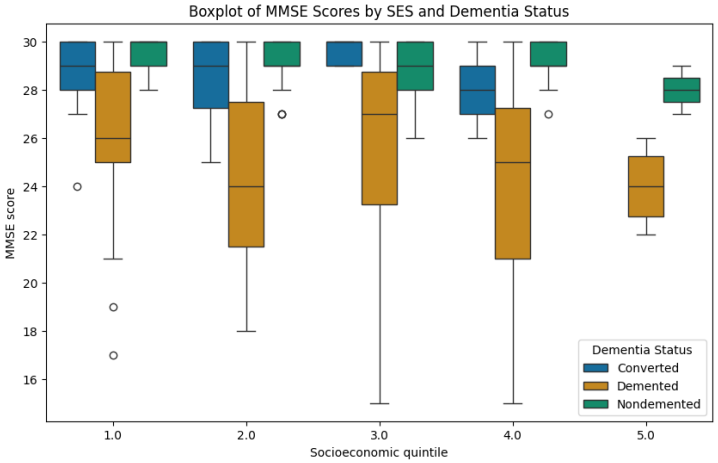

Figure 2

In this exploratory boxplot, we see that the particular relationships between SES and MMSE scores for the three dementia statuses are unclear. So, we proceeded with the between-subjects test to learn more. For this research question, MMSE score was the Dependent Variable, while SES and Group were Independent Variables and Subject ID was the random effect. Using a Shapiro-Wilk test on the MMSE data, we found that the data was not normally distributed, as the test statistic was 0.91 and the p-value was <0.0001 – so, we rejected the Null Hypothesis that the data was normally distributed. However, using a Box Cox transformation, we found that the transformed MMSE data was closer to normally distributed, and so we moved forward with the mixed-effects ANOVA from there. The results were as follows:

	Coef.	Std. Err.	z	P > z	[0.025	0.975]
Intercept	29.246	0.734	39.839	0.000	27.807	30.685
Group [T.Demented]	-3.869	0.743	-5.206	0.000	-5.326	-2.412
Group [T.Nondemented]	0.536	0.716	0.749	0.454	-0.867	1.940
SES	-0.259	0.185	-1.403	0.161	-0.621	0.103
Group Var	4.433	0.605				

There are several findings we extract from the results above. The Demented group had significantly lower MMSE scores than the Nondemented group, which aligns with what we would expect considering what MMSE is designed to measure. The Nondemented group, however, did not statistically differ in mean. Finally, considering SES, there is a slightly negative relationship between SES and MMSE (-0.259), but with a large p-value of 0.161, significantly higher than our alpha of 0.05. Therefore, we fail to reject the Null Hypothesis that SES has a significant effect on MMSE given the groups features in this dataset. Our response for Research

Question 1 is that we did confirm that Demented individuals do show statistically significantly lower MMSE scores than otherwise, but there was no further statistically significant conclusion to draw from our mixed-effects ANOVA.

V. Research Question #2: Within-Subjects

For this research question, we performed a mixed-effects ANOVA using the MMSE values from Research Question 1 as the Dependent Variable, with Visits (first or second) as the control and SES as the Independent Variable. The mixed-effects ANOVA performed produced these results:

	Coef.	Std. Err.	z	P > z	[0.025	0.975]
Intercept	28.824	0.942	30.592	0.000	26.977	30.670
SES	-0.248	0.345	-0.721	0.471	-0.924	0.427
Visit	-0.004	0.460	-0.009	0.993	-0.906	0.898
SES:Visit	-0.230	0.168	-1.366	0.172	-0.560	0.100
Group Var	8.790	1.086				

From these results, we find that, here too, there are no statistically significant results. The p-values for the coefficients presented in the chart are all significantly above the alpha of 0.05. Therefore, we fail to reject the Null Hypotheses that there are no significant relationships between Visit, MMSE score, and SES.

This statistically insignificant conclusions for both Research Question 1 and Research Question 2 are hardly exciting, but also not disappointing, and nonetheless beckon for further research: the interactions between SES and health states, and the relationship between health states and the tools used to measure those health states, are complex – with a larger dataset, more conclusions could be drawn for these questions.

V. References

Alzheimer's Disease International (n.d.). Dementia facts & figures. Alzheimer's Disease International. https://www.alzint.org/about/dementia-facts-figures/

Alzheimer's Disease International (2019). World Alzheimer report 2019: attitudes to dementia. Alzheimer's Disease International. https://www.alzint.org/u/WorldAlzheimerReport2019.pdf

Vliet, D., Vugt, M. E., Bakker, C., Koopmans, R., and Verhey, Frans (April 20, 2010). Impact of early onset dementia on caregivers: a review. *International Journal of Geriatric Psychiatry* 25(11), 1091-1100. https://doi-org.myaccess.library.utoronto.ca/10.1002/gps.2439